Structure of Dietary Measurement Error: Results of the OPEN Biomarker Study

Victor Kipnis[1], Amy F. Subar[2], Douglas Midthune[1], Laurence S. Freedman[3,4], Rachel Ballard-Barbash[2], Richard P. Troiano[2], Sheila Bingham[5], Dale A. Schoeller[6], Arthur Schatzkin[7], and Raymond J. Carroll[8]

[1] Biometry Research Group, Division of Cancer Prevention, National Cancer Institute, Bethesda, MD.
[2] Applied Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD.
[3] Department of Mathematics, Statistics and Computer Science, Bar Ilan University, Ramat Gan, Israel.
[4] Gertner Institute for Epidemiology and Health Policy Research, Tel Hashomer, Israel.
[5] Medical Research Council, Dunn Human Nutrition Unit, Cambridge, United Kingdom.
[6] Department of Nutritional Sciences, University of Wisconsin, Madison, WI.
[7] Nutritional Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD.
[8] Department of Statistics, Texas A&M University, College Station, TX.

Multiple-day food records or 24-hour dietary recalls (24HRs) are commonly used as "reference" instruments to calibrate food frequency questionnaires (FFQs) and to adjust findings from nutritional epidemiologic studies for measurement error. Correct adjustment requires that the errors in the adopted reference instrument be independent of those in the FFQ and of true intake. The authors report data from the Observing Protein and Energy Nutrition (OPEN) Study, conducted from September 1999 to March 2000, in which valid reference biomarkers for energy (doubly labeled water) and protein (urinary nitrogen), together with a FFQ and 24HR, were observed in 484 healthy volunteers from Montgomery County, Maryland. Accounting for the reference biomarkers, the data suggest that the FFQ leads to severe attenuation in estimated disease relative risks for absolute protein or energy intake (a true relative risk of 2 would appear as 1.1 or smaller). For protein adjusted for energy intake by using either nutrient density or nutrient residuals, the attenuation is less severe (a relative risk of 2 would appear as approximately 1.3), lending weight to the use of energy adjustment. Using the 24HR as a reference instrument can seriously underestimate true attenuation (up to 60% for energy-adjusted protein). Results suggest that the interpretation of findings from FFQ-based epidemiologic studies of diet-disease associations needs to be reevaluated.

bias (epidemiology); biological markers; diet; energy intake; epidemiologic methods; nutrition assessment; questionnaires; reference values

Abbreviations: DLW, doubly labeled water; FFQ, food frequency questionnaire; OPEN, Observing Protein and Energy Nutrition; 24HR, 24-hour dietary recall.

Much of the recent literature on the relation between diet and cancer has been based on analytic epidemiologic studies using food frequency questionnaires (FFQs). A number of large prospective studies of this kind have failed to find a consistent relation between dietary components (such as fat, fiber, and fruits and vegetables) and cancers of the breast, colon, or rectum (1–3), which may be explained by a true lack of diet-cancer associations or, alternatively, by serious methodological limitations of the studies themselves, especially due to FFQ measurement error.

Over the years, investigators have recognized that the reported values from FFQs are subject to substantial error, both systematic and random, that can profoundly affect the design, analysis, and interpretation of nutritional epidemiologic studies (4–6). Dietary measurement error often attenuates (biases toward one) the estimates of disease relative

risks and reduces statistical power to detect their significance. Therefore, an important relation between diet and disease may be obscured.

This problem has prompted researchers involved in large epidemiologic investigations to integrate calibration substudies that include a more intensive, but presumably more accurate, reference method, typically multiple-day food records (7) or multiple 24-hour dietary recalls (24HRs) (8). Comparing reference measurements with those from the FFQ enables adjustment for attenuation by using the regression calibration approach (7). However, the correct application of this approach requires that the adopted reference instrument satisfy two critical conditions. Although it may be imperfect and contain measurement error, this error should be independent of 1) true intake and 2) error in the FFQ (9). Throughout this paper, we take these two conditions as requirements for a valid reference instrument.

A great deal of accumulated evidence suggests that common dietary report reference instruments are unlikely to meet these requirements. Studies with the few biomarkers of dietary intake that do qualify as valid reference measurements ("reference" biomarkers), such as doubly labeled water (DLW) for total energy expenditure and urinary nitrogen for protein intake, demonstrate serious systematic biases in all dietary report instruments that may be potentially related (10–16). This has led to proposals for new models of dietary measurement error that might explain why the large prospective studies fail to find a relation between diet and cancer, even were an important relation to exist (9, 17, 18).

For example, Kipnis et al. (9) considered two potential systematic components of dietary measurement error. The first component reflects correlation between error and true intake ("intake-related" bias). The second component ("person-specific" bias) is independent of true intake and represents the difference between total within-person bias and its intake-related component. The existence of person-specific biases was proposed in all dietary report instruments, and a sensitivity analysis demonstrated that correlation between person-specific biases in the FFQ and the reference instrument, if ignored, could lead to serious underestimation of the degree of attenuation in a conventional calibration study. In a subsequent paper, Kipnis et al. (18) provided empirical evidence directly supporting their hypothesis, based on the results from a validation study that included the urinary nitrogen reference biomarker for protein intake. Moreover, based on the urinary nitrogen data, the measurement error model was extended to also include intake-related bias in dietary report reference instruments and was shown to fit the data statistically significantly better than other proposed models.

In this paper, we take this further by analyzing data from the Observing Protein and Energy Nutrition (OPEN) Study that included reference biomarkers for protein (urinary nitrogen) and energy (DLW) intakes, together with a FFQ and a 24HR. This study enabled us to evaluate not only absolute protein intake but also total energy and energy-adjusted protein intakes (19). We were therefore able to investigate the conjecture that energy adjustment substantially reduces measurement error in reported intake and that remaining error can be reliably corrected for by the common approach (20).

MATERIALS AND METHODS

Effect of measurement error

The effects of dietary measurement error on the estimation of disease risks are well known (9). The most important concept is that of attenuation. Consider the disease model

$$R(D|T) = \alpha_0 + \alpha_1 T, \tag{1}$$

where $R(D|T)$ denotes the risk of disease D on an appropriate scale (e.g., logistic) and T is true habitual intake of a given nutrient, also measured on an appropriate scale. The slope α_1 represents an association between the nutrient intake and disease (e.g., log relative risk). In practice, FFQ-reported intake Q is used instead of unknown true intake T. We assume throughout that dietary measurement error is nondifferential with respect to disease D; that is, reported intake contributes no additional information about disease risk beyond that provided by true intake. To an excellent approximation, fitting model 1 to reported intake leads to estimating not the true risk parameter α_1 but the product $\tilde{\alpha}_1 = \lambda_1 \alpha_1$ of α_1 and the slope λ_1 in the linear regression calibration model, $T = \lambda_0 + \lambda_1 Q + \xi$, where ξ denotes random error.

In nutritional studies, the value of λ_1 is usually between 0 and 1 (21), so dietary measurement error leads to underestimation of the true risk parameter. This underestimation is called attenuation, and λ_1 is called the attenuation factor. Values of λ_1 closer to zero lead to more serious underestimation of risk. For example, a true relative risk of 2 would appear as $2^{0.4} = 1.32$ if the attenuation factor were 0.4 and as $2^{0.2} = 1.15$ if the attenuation factor were 0.2.

Measurement error also leads to loss of statistical power for testing disease-exposure associations. Approximately, the sample size required to reach the desired statistical power to detect a given risk is proportional to $N \propto 1/\{\rho^2(Q,T)\sigma_T^2\} = 1/\{\lambda_1^2\sigma_Q^2\}$, where $\rho(Q,T)$ is the correlation between the reported and true intakes and σ_Q^2 and σ_T^2 are the between-person variances of the reported and true intakes, respectively (21). In particular, for a given FFQ, the required sample size is inversely proportional to the squared attenuation factor, λ_1^2. For example, if the true attenuation factor were 0.2, the sample size, calculated by assuming that $\lambda_1 = 0.4$, should be multiplied by $0.4^2/0.2^2 = 4$ to achieve the nominal power.

Estimation of the attenuation factor

Estimation of the attenuation factor λ_1 requires collecting additional reference measurements to compare with the FFQ in the calibration substudy (9). The common approach in nutritional epidemiology uses a more intensive dietary report method as the reference instrument, assuming that it is unbiased at the individual level and that its errors are independent of those in the FFQ (7). In this paper, we contrast this model with the measurement error model of Kipnis et al. (18) that specifies the same general error structure in the dietary

report reference instrument (F) as the one for the FFQ (Q). To be fully identifiable, the model requires data from a reference biomarker. The model is specified as

$$Q_{ij} = \mu_{Qj} + \beta_{Q0} + \beta_{Q1}T_i + r_i + \varepsilon_{ij}$$
$$F_{ij} = \mu_{Fj} + \beta_{F0} + \beta_{F1}T_i + s_i + u_{ij} \qquad (2)$$
$$M_{ij} = \mu_{Mj} + T_i + \upsilon_{ij},$$

where μ_{Qj}, μ_{Fj}, and μ_{Mj} are time-specific group intercepts for the FFQ, 24HR, and biomarker, respectively, which sum to zero over j; β_{Q0} and β_{F0} are the overall group intercepts for the FFQ and 24HR; β_{Q1} and β_{F1} are the slopes reflecting intake-related bias for the FFQ and 24HR; r_i and s_i are person-specific biases for the FFQ and 24HR that are independent of true intake T_i, have means zero, variances σ_r^2 and σ_s^2, respectively, and are correlated with the correlation coefficient ρ_{rs}; and ε_{ij}, u_{ij}, and υ_{ij} are within-person random errors for the FFQ, 24HR, and biomarker, with means zero and variances σ_ε^2, σ_u^2, and σ_υ^2, respectively, that are assumed to be independent of each other and of other terms in the model, except that "within-pair" errors (ε_{ij}, u_{ij}), (ε_{ij}, υ_{ij}), and (u_{ij}, υ_{ij}) are allowed to be correlated, if the corresponding measurements are taken contemporaneously.

In the presence of the reference biomarker, model 2 does not require an instrument F to estimate the error components in the FFQ. However, its inclusion enables us to additionally analyze the error structure of the dietary report reference instrument and its relation to that in the FFQ.

The common model may be obtained from model 2 by ignoring information from the reference biomarker and assuming that the dietary report instrument F contains no intake-related bias ($\beta_{F1} = 1$) or person-specific bias ($\sigma_s^2 = 0$). We use the following general form of this model:

$$Q_{ij} = \mu_{Qj} + \beta_{Q0} + \beta_{Q1}T_i + r_i + \varepsilon_{ij},$$
$$F_{ij} = \mu_{Fj} + T_i + u_{ij}. \qquad (3)$$

When the model parameters are used, the attenuation factor is expressed as

$$\lambda_1 = \frac{\text{cov}(T, Q)}{\text{var}(Q)} = \frac{\beta_{Q1}}{\beta_{Q1}^2 + \sigma_r^2/\sigma_T^2 + \sigma_\varepsilon^2/\sigma_T^2}, \qquad (4)$$

and the correlation of the FFQ and true intake is given by

$$\rho_{Q,T} = \frac{\text{cov}(T, Q)}{\sqrt{\text{var}(T)\text{var}(Q)}} = \frac{\beta_{Q1}}{\sqrt{\beta_{Q1}^2 + \sigma_r^2/\sigma_T^2 + \sigma_\varepsilon^2/\sigma_T^2}}. \qquad (5)$$

Both are estimated by replacing the parameters by their estimates based on the corresponding model 2 or 3. Doing so is essentially equivalent to adjusting for random measurement error in the adopted reference instrument.

The OPEN data

The OPEN Study was conducted by the National Cancer Institute from September 1999 to March 2000. The recruitment procedure, subject characteristics, and detailed study conduct are described in the companion paper in this issue of the *Journal* (22). Briefly, 261 male and 223 female participants aged 40–69 years were healthy volunteers from Montgomery County, Maryland. Each participant was asked to complete a FFQ and a 24HR on two occasions. The FFQ was completed within 2 weeks of visit 1 and then approximately 3 months later, within a few weeks of visit 3. The 24HR was completed at visit 1 and then approximately 3 months later at visit 3. Participants received their DLW dose at visit 1 and returned 2 weeks later (visit 2) to complete the DLW assessment. In addition, repeat DLW measurements were collected from 14 male and 11 female volunteers who received their second DLW dose at the end of visit 2 and returned 2 weeks later to complete their DLW assessment. Participants provided two 24-hour urine collections during the 2-week period between visit 1 and visit 2, verified for completeness by using the PABAcheck method (23). Since approximately 81 percent of nitrogen intake is excreted through the urine (18) and nitrogen constitutes 16 percent of protein, the urinary nitrogen values were adjusted, by dividing by 0.81 and multiplying by 6.25, to estimate individual protein intake.

The adopted FFQ was the Diet History Questionnaire, developed and evaluated at the National Cancer Institute (24–28). The 24HR was a highly standardized version using the five-pass method, developed by the US Department of Agriculture for use in national dietary surveillance (29).

Statistical analysis

Throughout, we applied the logarithmic transformation to energy and protein to make measurement error in the DLW and urinary nitrogen biomarkers additive and homoscedastic and to better approximate normality. In addition to total energy and protein, the reference biomarkers in the OPEN Study enabled us to evaluate dietary measurement error for energy-adjusted protein intake. Because modeling relations between disease and multiple covariates measured with error is beyond the scope of this paper, we assumed that model 1 included only energy-adjusted exposure and that energy was not related to disease. We used two energy adjustment methods: nutrient density and nutrient residual (19). Protein density was calculated as the percentage of energy from protein sources and was then log transformed. The protein residual was calculated from the linear regression of protein on energy intake on the log scale. Both protein density and residual were calculated for each instrument by using the protein and energy intakes as measured by this instrument. The convention used for dealing with biomarker-based derived measures is explained in the Appendix.

For all dietary variables, we excluded extreme outlying values that fell outside the interval given by the 25th percentile minus twice the interquartile range to the 75th percentile plus twice the interquartile range. For each variable and each

TABLE 1. Estimated attenuation factor λ_1 and correlation $\rho(Q, T)$ of food frequency questionnaire-reported intake (Q) and true intake (T)* in the Observing Protein and Energy Nutrition Study, Maryland, September 1999–March 2000

Nutrient	Gender	Model	Attenuation factor λ_1	Correlation $\rho(Q,T)$
Energy	Male	Biomarker based†	0.080 (0.025)‡	0.199 (0.061)
		24HR based§	0.230 (0.037)	0.437 (0.065)
	Female	Biomarker based	0.039 (0.028)	0.098 (0.069)
		24HR based	0.128 (0.044)	0.261 (0.088)
Protein	Male	Biomarker based	0.156 (0.034)	0.323 (0.067)
		24HR based	0.177 (0.043)	0.312 (0.074)
	Female	Biomarker based	0.137 (0.041)	0.298 (0.088)
		24HR based	0.158 (0.046)	0.334 (0.098)
Protein density	Male	Biomarker based	0.404 (0.066)	0.431 (0.063)
		24HR based	0.409 (0.066)	0.497 (0.077)
	Female	Biomarker based	0.316 (0.084)	0.346 (0.087)
		24HR based	0.501 (0.060)	0.778 (0.117)

* As estimated by the model accounting for the reference biomarker of intake or the common model accounting only for the 24-hour recall (24HR) as reference measurements.

† Defined as model 2 in the text.

‡ Numbers in parentheses, standard error.

§ Defined as model 3 in the text.

instrument, no more than six outlying values for men and four for women were excluded from the analyses.

The estimates of the model parameters and their standard errors were obtained by using the method of maximum likelihood under the assumption of normality of the random terms in the models. Standard errors were checked for accuracy by using the bootstrap method. Comparisons of correlated parameters (such as attenuation factors estimated by two models) were performed by comparing the ratios of their differences to the standard errors of the differences calculated by the bootstrap method with the standardized normal distribution.

RESULTS

The descriptive statistics for measurements taken by using the different instruments are provided in the companion paper (22). For energy-adjusted protein, the results for only nutrient density are shown since the results for nutrient residual were similar.

Attenuation and correlation with true intake

Table 1 displays the estimates of the attenuation factor λ_1 and correlation $\rho(Q, T)$ between the FFQ and true usual intake resulting from applying models 2 and 3 to energy, protein, and energy-adjusted protein. The table contrasts the estimated values when the common approach versus the biomarker-based model was used.

Absolute intakes. The biomarker-based attenuation factors were distressingly close to zero. For example, for women, the attenuation factors for energy and protein were 0.039 and 0.137, respectively. The attenuation factors esti-

mated by using the common approach were substantially higher (underestimating the corresponding attenuation) for energy at 0.128 ($p = 0.05$ when compared with the biomarker-based attenuation) and somewhat higher for protein at 0.158 ($p = 0.73$). Results for men showed a similar pattern, with the attenuation factor being statistically significantly overestimated ($p < 0.001$) when the common approach for energy was used.

The correlations between the FFQ and true intake were also very low. The biomarker-based correlations for energy and protein intakes for women were 0.098 and 0.298, respectively, while the common approach overestimated correlations at 0.261 ($p = 0.10$) and 0.334 ($p = 0.81$). For men, the correlation estimated by using the common approach was statistically significantly biased upward ($p < 0.001$) for energy.

Energy-adjusted intakes. For energy-adjusted intakes, the attenuation factors were somewhat higher (attenuation was lower) than for absolute intakes. For example, for women, the biomarker-based estimate for protein density was 0.316 compared with 0.137 for protein ($p = 0.10$). Results for men showed a similar pattern, with the highly statistically significant difference in attenuation between absolute and energy-adjusted protein intakes ($p < 0.001$).

The attenuation factor estimated by using the common approach for women again appeared substantially more optimistic than the biomarker-based estimate at 0.501 versus 0.316 ($p = 0.10$) for protein density. For men, however, no marked difference was found between the attenuation factors estimated by using the two models. Correlations between FFQ and true intake for energy-adjusted protein displayed the same pattern as those for attenuation factors.

TABLE 2. Variance of true intake and parameters of dietary measurement error in the food frequency questionnaire and 24-hour dietary recall,* the Observing Protein and Energy Nutrition Study, Maryland, September 1999–March 2000

Nutrient	Gender	Model	Variance of true intake $\sigma^2_T \times 10^2$	Slope in regression of FFQ†-reported on true intake β_{Q1}	Slope in regression of 24HR-reported on true intake β_{F1}	Variance of person-specific bias in FFQ $\sigma^2_r \times 10^2$	Variance of person-specific bias in 24HR $\sigma^2_s \times 10^2$	Correlation of person-specific biases in FFQ and 24HR $\rho_{r,s}$	Variance of within-person error in FFQ $\sigma^2_\epsilon \times 10^2$	Variance of within-person error in 24HR $\sigma^2_u \times 10^2$
Energy	Male	Biomarker based‡	2.6 (0.27)§	0.49 (0.15)	0.66 (0.10)	12.2 (1.2)	3.2 (0.61)	0.45 (0.08)	3.2 (0.28)	5.3 (0.48)
		24HR based¶	4.4 0.68	0.83 (0.15)	1	9.7 (1.2)	0		3.2 (0.28)	5.3 (0.47)
	Female	Biomarker based	2.4 (0.29)	0.24 (0.17)	0.46 (0.13)	11.2 (1.3)	3.2 (0.78)	0.28 (0.11)	3.9 (0.37)	7.9 (0.75)
		24HR based	3.7 (0.81)	0.53 (0.20)	1	10.3 (1.3)	0		3.9 (0.37)	7.9 (0.75)
Protein	Male	Biomarker based	4.4 (0.57)	0.67 (0.15)	0.70 (0.11)	13.3 (1.4)	3.9 (0.94)	0.18 (0.10)	3.7 (0.33)	9.3 (0.82)
		24HR based	6.1 (1.0)	0.55 (0.14)	1	13.5 (1.5)	0		3.7 (0.33)	9.3 (0.82)
	Female	Biomarker based	3.7 (0.71)	0.65 (0.21)	0.60 (0.16)	11.0 (1.5)	2.6 (1.1)	0.24 (0.15)	4.8 (0.46)	12.0 (1.2)
		24HR based	3.9 (1.2)	0.70 (0.25)	1	10.7 (1.6)	0		4.8 (0.46)	12.0 (1.2)
Protein density	Male	Biomarker based	3.1 (0.47)	0.46 (0.08)	0.62 (0.11)	1.6 (0.25)	1.2 (0.50)	0.40 (0.15)	1.2 (0.11)	5.8 (0.51)
		24HR based	2.4 (0.53)	0.60 (0.13)	1	1.4 (0.30)	0		1.2 (0.11)	5.8 (0.51)
	Female	Biomarker based	3.5 (0.72)	0.38 (0.11)	0.39 (0.13)	2.3 (0.36)	1.2 (0.60)	0.94 (0.19)	1.4 (0.13)	6.8 (0.65)
		24HR based	1.7 (0.59)	1.2 (0.36)	1	0.30 (0.76)	0		1.4 (0.13)	6.8 (0.65)

* As estimated by the model accounting for the reference biomarker of intake or the common model accounting only for the 24-hour dietary recall (24HR) as reference measurements

† FFQ, food frequency questionnaire.

‡ Defined as model 2 in the text.

§ Numbers in parentheses, standard error.

¶ Defined as model 3 in the text.

Error structure of the FFQ and 24HR

Intake-related bias. Table 2 demonstrates across-the-board intake-related bias in both FFQ and 24HR measurements. All biomarker-based estimates of slopes β_{Q1} and β_{F1} were substantially smaller than the desired value of 1.0, leading to the flattened slope phenomenon. If anything, energy adjustment seemed to make this phenomenon even more pronounced. The flattened slope in the FFQ estimated by using the common approach is often not seen as clearly. For example, for males, the DLW-based estimate of β_{Q1} for energy intake was 0.49, but the common estimate was 0.83.

Person-specific bias. Table 2 also demonstrates the existence and importance of person-specific biases in reported intakes from the FFQ and 24HR. Compared with the true between-person variance σ^2_T, the person-specific biases σ^2_s and σ^2_r were quite dominant for absolute intakes. For example, for females reporting protein intake, the FFQ person-specific bias variance was 0.110 and the 24HR person-specific bias variance was 0.026, quite large compared with the variance of true intake (0.037). Energy adjustment considerably reduced person-specific biases. Continuing with the example above, for protein density, this variance was reduced from 0.110 to 0.023 for the FFQ and from 0.026 to 0.012 for the 24HR, while the variance of true intake remained practically the same (0.035). However, even for energy-adjusted intakes, person-specific biases were still substantial and highly significantly different from zero.

Table 2 also demonstrates substantial positive correlation $\rho_{r,s}$ between person-specific biases in the FFQ and 24HR. The correlation increased after energy adjustment, especially for women.

Within-person random error. For absolute intakes, within-person random variation σ^2_ϵ in the FFQ was of the same magnitude as between-person variation σ^2_T of true intake. Similar to person-specific bias, it was considerably reduced by energy adjustment. As expected because of day-to-day variation in intake, within-person random variation σ^2_u in the 24HR was substantially greater. Interestingly, relative to variation of true intake, it was only moderately reduced by energy adjustment. In all cases considered, within-person random errors were not statistically significantly correlated across instruments.

"Nonprotein" intake. Using the measurements for protein and energy on each instrument, we also evaluated dietary measurement error for nonprotein-energy-contributed nutrients ("nonprotein" for short), for both absolute nonprotein and energy-adjusted nonprotein intakes. The results for absolute nonprotein intake were similar to the results for energy, and the results for energy-adjusted nonprotein were similar to the results for energy-adjusted protein.

DISCUSSION

In this paper, we focused mostly on the attenuation factor because it directly affects the observed relative risks and the sample size necessary to detect diet-disease associations in epidemiologic studies. The critical requirement for our results that the adopted biomarkers represent valid reference instruments, that is, their errors are unrelated to true intakes and errors in dietary report instruments, is supported by accumulated evidence for both the adjusted urinary nitrogen (18) and DLW (30). The OPEN Study yielded the following conclusions.

First, the impact of FFQ measurement error on total energy and absolute protein intakes was severe and in agreement with the findings of Kipnis et al. (18) for protein intake. Attenuation factors were vexingly close to zero, as were the correlations with true intake.

Second, the impact of measurement error seemed less severe after energy adjustment. As follows from expression 4, the attenuation factor is inversely proportional to the variances of both person-specific bias and within-person random error relative to between-person variation of true intake. Since these relative variances decreased substantially after energy adjustment (table 2) because of correlated errors in reporting protein and energy, energy-adjusted protein was less affected by measurement error compared with absolute protein intake. However, the estimated attenuation factors for energy-adjusted intakes were in the range 0.32–0.41 (table 1), indicating that measurement error still remained an important problem.

Third, the 24HR was seriously flawed, suffering from intake-related bias and from person-specific bias that was correlated with person-specific bias in the FFQ. As a result, it violated both requirements for a valid reference instrument and in most cases substantially misrepresented the impact of measurement error in the FFQ. As follows from formula A1 in the Appendix, bias in the attenuation factor λ_F calculated by using the common approach depends on the sum of the values for slope β_{F1} and expression

$$\rho_{r,s}\sqrt{(\sigma_r^2/\sigma_T^2)(\sigma_s^2/\sigma_T^2)}/\beta_{Q1}.$$

Table 2 reveals that, for absolute intakes, the relative variances of person-specific biases in the FFQ and 24HR and the correlation between them were sufficiently large to override the small values of β_{F1} and to raise λ_F above the true attenuation factor λ_1. The same remained true for energy-adjusted protein in women, where the effect of reduced person-specific biases was compensated for by the increased correlation between them. As a result, the 24HR underestimated true attenuation. On the other hand, for energy-adjusted protein in men, the two effects essentially cancelled each other, demonstrating that a flawed reference instrument may sometimes produce a good estimate.

Our results are in line with previous data presented on protein intake. For women in the British Medical Research Council study (18), the urinary-nitrogen-based attenuation factor for protein was 0.187, while the common approach based on a 4-day weighed food record produced an overly optimistic estimate of 0.282. The former is slightly larger than the 0.137 obtained in the OPEN Study, while the latter is noticeably more optimistic than our 24HR-based estimate of 0.158 ($p = 0.08$). The correlations of FFQ with true intake were 0.284 (urinary nitrogen based) and 0.432 (record based) compared with our values of 0.298 (urinary nitrogen based) and 0.334 (24HR based), respectively. Neither difference approaches statistical significance.

An important consideration is whether our results could be affected by the fact that biomarkers in the OPEN Study were collected mostly over one season. We analyzed 24HRs taken in different seasons in cross-sectional national survey data (Continuing Survey of Food Intakes by Individuals 1994–1996) by region and gender, and we found no seasonal fluctuations in energy or protein intakes. However, if seasonality were to exist, it would affect only the estimated mean usual intake and would not change the higher-order parameters presented in tables 1 and 2.

Since DLW measures total energy expenditure, it would be important to adjust the data for long-term weight change to enable DLW to truly represent usual energy intake. Doing so over the 2-week DLW period may introduce only more random error, however, since only a small amount of within-person week-to-week fluctuations in energy balance can be explained by contemporary changes in weight (31). Even using the 3-month OPEN Study period may not adequately represent long-term weight changes, especially given protocol differences in fasting conditions between the first and last visits (22). Nevertheless, when we adjusted individual DLW measurements for the weight change over either the 2-week or 3-month period, the results did not change materially for either absolute or energy-adjusted nutrients.

Recently, Willett (20) suggested that any evaluation of a FFQ would be invalid unless heterogeneity in the study population due to gender, age, and body size was adjusted for. To address this issue, we performed further analyses that included age in 5-year groups and the logarithm of body mass index as covariates in the models. The results did not change substantially except for energy in women; the attenuation factor and correlation of the FFQ with true intake became even closer to zero.

Our results have important implications for nutritional epidemiology. First, they question the ability of FFQs to detect diet-disease associations for absolute nutrient intakes. While some journals have recently required that energy adjustment be used in the analysis of nutrient-disease associations, the practice has been controversial (32, 33). Our data clearly document failure of the FFQ to provide a sufficiently accurate report of absolute protein, nonprotein, and energy intakes to enable detection of their moderate associations with disease. For example, with the attenuation factors of 0.08 for energy intake for males and 0.04 for females, a true relative risk of 2.0 would appear as 1.06 and 1.03, respectively, using the FFQ data. Needless to say, such small relative risks are not detectable in epidemiologic studies since their signal is smaller than the noise caused by confounders. It is plausible that similarly small attenuation factors would be found for many other nutrients, although it would require a suitable reference biomarker for each nutrient to confirm this possibility.

Second, it appears that FFQ-based energy-adjusted nutrient intakes may just be sufficiently accurate to use in large cohort studies to detect moderate diet-disease associations; a relative risk of 2.0 would appear close to 1.3, which could be at the limits of detection. The benefits of adjusting for energy intake have been discussed previously at the general level (19, 32). Our conclusion is necessarily a qualified one, since our study was restricted to energy-adjusted protein and nonprotein intakes. There is no guarantee that the results will be as favorable for nonprotein components such as energy-adjusted fat intake. Even less could be speculated about the effect of energy adjustment for non-energy-contributing nutrients. Nevertheless, until further evidence

becomes available on other nutrients, use of energy-adjusted intakes seems the best working approach for nutritional epidemiology, at least under the assumption that energy is not related to disease. Note, however, that biomarker-based attenuation factors for energy-adjusted protein intake are between 0.32 and 0.41, indicating that measurement error has a substantial negative impact on the statistical power of observational epidemiologic studies.

Third, our results throw into question use of the 24HR as a reference instrument for validation/calibration studies. In the OPEN Study, such use substantially overestimated performance of the FFQ for absolute intakes of energy and nonprotein. The results also cast some doubt on the performance of the 24HR as a reference for energy-adjusted intakes. For example, for protein density in women (table 1), the biomarker-based attenuation factor was estimated as 0.3 compared with the 24HR-based estimate of 0.5. Use of the latter would lead to underestimation of the required sample size by a factor of $2.8 = 0.5^2/0.3^2$, with profound effects on the power to detect diet-disease associations.

The OPEN Study provides solid evidence of measurement errors in a FFQ as they pertain to energy intake and both absolute and energy-adjusted protein and nonprotein intakes. Further studies of a similar design are needed to confirm our results, especially to clarify whether 24HRs or multiple-day food records can be used reliably as reference instruments in validation/calibration studies, at least for energy-adjusted intakes. Unfortunately, few dietary biomarkers qualify as valid reference instruments; that is, they have errors unrelated to true intakes and errors in dietary report instruments. Most other biomarkers, such as vitamin C or beta-carotene, measure concentrations of related constituents for which the quantitative relation to dietary intake is unknown and depends on individual characteristics (e.g., concomitant intake of other nutrients, obesity, or smoking habits) (34). Therefore, such concentration-based biomarkers cannot provide valid reference measurements and at best can serve only as correlates of intake. Further work should explore whether a combination of data from dietary report and biomarker measurements for energy or protein can be used to assess dietary exposure variables for which no reference biomarkers exist.

ACKNOWLEDGMENTS

REFERENCES

1. Hunter DJ, Spiegelman D, Adami HO, et al. Cohort studies of fat intake and the risk of breast cancer—a pooled analysis. N Engl J Med 1996;334:356–61.
2. Fuchs CS, Giovannucci EL, Colditz GA, et al. Dietary fiber and the risk of colorectal cancer and adenoma in women. N Engl J Med 1999;340:169–76.
3. Michels KB, Giovannucci E, Joshipura KJ, et al. Prospective study of fruit and vegetable consumption and incidence of colon and rectal cancers. J Natl Cancer Inst 2000;92:1740–52.
4. Beaton GH, Milner J, Corey P, et al. Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. Am J Clin Nutr 1979;32:2546–59.
5. Freudenheim JL, Marshall JR. The problem of profound mismeasurement and the power of epidemiologic studies of diet and cancer. Nutr Cancer 1988;11:243–50.
6. Freedman LS, Schatzkin A, Wax Y. The impact of dietary measurement error on planning a sample size required in a cohort study. Am J Epidemiol 1990;132:1185–95.
7. Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. Stat Med 1989;8:1051–69.
8. Kaaks R, Riboli E. Validation and calibration of dietary intake measurements in the EPIC project: methodological considerations. Int J Epidemiol 1997;26(suppl):S15–25.
9. Kipnis V, Carroll RJ, Freedman LS, et al. Implications of a new dietary measurement error model for estimation of relative risk: application to four calibration studies. Am J Epidemiol 1999;150:642–51.
10. Bandini LG, Schoeller DA, Cyr HN, et al. Validity of reported energy intake in obese and nonobese adolescents. Am J Clin Nutr 1990;52:421–5.
11. Livingstone MBE, Prentice AM, Strain JJ, et al. Accuracy of weighed dietary records in studies of diet and health. BMJ 1990;300:708–12.
12. Heitmann BL. The influence of fatness, weight change, slimming history and other lifestyle variables on diet reporting in Danish men and women aged 35–65 years. Int J Obes 1993;17:329–36.
13. Heitmann BL, Lissner L. Dietary underreporting by obese individuals—is it specific or non-specific? BMJ 1995;311:986–9.
14. Martin LJ, Su W, Jones PJ, et al. Comparison of energy intakes determined by food records and doubly labeled water in women participating in a dietary-intervention trial. Am J Clin Nutr 1996;63:483–90.
15. Sawaya AL, Tucker K, Tsay R, et al. Evaluation of four methods for determining energy intake in young and older women: comparison with doubly labeled water measurements of total energy expenditure. Am J Clin Nutr 1996;63:491–9.
16. Black AE, Bingham SA, Johansson G, et al. Validation of dietary intakes of protein and energy against 24 urinary N and DLW energy expenditure in middle-aged women, retired men and post-obese subjects: comparisons with validation against presumed energy requirements. Eur J Clin Nutr 1997;51:405–13.
17. Prentice R. Measurement error and results from analytic epidemiology: dietary fat and breast cancer. J Natl Cancer Inst 1996;88:1738–47.
18. Kipnis V, Midthune D, Freedman LS, et al. Empirical evidence of correlated biases in dietary assessment instruments and its implications. Am J Epidemiol 2001;153:394–403.
19. Willett WC. Nutritional epidemiology. Chapter 5. New York, NY: Oxford University Press, 1990.
20. Willett W. Commentary: dietary diaries versus food frequency questionnaires—a case of undigestible data. Int J Epidemiol 2001;30:317–19.
21. Kaaks R, Riboli E, van Staveren W. Calibration of dietary intake measurements in prospective cohort studies. Am J Epidemiol 1995;142:548–56.
22. Subar AF, Kipnis V, Troiano RP, et al. Using intake biomarkers

to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN Study. Am J Epidemiol 2003;158:1–13.

23. Bingham SA, Cummings JH. Urine nitrogen as an independent validatory measure of dietary intake: a study of nitrogen balance in individuals consuming their normal diet. Am J Clin Nutr 1985;42:1276–89.

24. Subar AF, Thompson FE, Smith AF, et al. Improving food frequency questionnaires: a qualitative approach using cognitive interviewing. J Am Diet Assoc 1995;95:781–8.

25. Subar AF, Midthune D, Kulldorff M, et al. Evaluation of alternative approaches to assign nutrient values to food groups in food frequency questionnaires. Am J Epidemiol 2000;152:279–86.

26. Subar AF, Ziegler RG, Thompson FE, et al. Is shorter always better? Relative importance of questionnaire length and cognitive ease on response rates and data quality for two dietary questionnaires. Am J Epidemiol 2001;153:404–9.

27. Subar AF, Thompson FE, Kipnis V, et al. Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires: the Eating at America's Table Study. Am J Epidemiol 2001;154:1089–99.

28. Thompson FE, Subar AF, Brown CC, et al. Cognitive research enhances accuracy of food frequency questionnaire reports: results of an experimental validation study. J Am Diet Assoc 2002;102:212–25.

29. Moshfegh AJ, Raper N, Ingwersen I, et al. An improved approach to 24-hour dietary recall methodology. Ann Nutr Metab 2001;45(suppl 1):156.

30. Schoeller DA. Measurement error of energy expenditure in free-living humans by using doubly labeled water. J Nutr 1988; 118:1278–89.

31. Edholm OG, Healy MJR, Wolfe HS, et al. Food intake and energy expenditure in army recruits. Br J Nutr 1970;24:1091–107.

32. Willett W, Howe GR, Kushi L. Adjustment for total energy intake in epidemiological studies. Am J Clin Nutr 1997; 65(suppl):1220S–8S.

33. Freedman LS, Kipnis V, Brown CC, et al. Comments on "Adjustment for total energy intake in epidemiological studies." Am J Clin Nutr 1997;65(suppl):1229S–31S.

34. Kaaks RJ. Biochemical markers as additional measurements in studies of the accuracy of dietary questionnaire measurements: conceptual issues. Am J Clin Nutr 1997;65(suppl):1232S–9S.

APPENDIX

Derived Reference Measures Based on the Observed Biomarkers

In the OPEN Study, replications of the DLW measurement were available for only a small sample of 25 persons (14 men and 11 women). This fact did not affect the results for total energy intake since the DLW measurements were remarkably consistent across replications. The coefficient of variation in the DLW measurements was only 5.1 percent, in

effect indicating that energy expenditure was measured with very little error.

However, a technical difficulty arose in the analysis of nonprotein and energy-adjusted nutrients. The error in the biomarker-based derived reference measures was almost entirely influenced by the error in the urinary nitrogen measurements, where the coefficient of variation was 17.6 percent. As a result, attempting to estimate the within-person variance of the derived reference measurements as a parameter in the model led to relatively large standard errors in the main analysis and to instability in the procedure for bootstrap calculations.

On the basis of these facts, in dealing with the derived reference measurements for nonprotein and energy-adjusted protein and nonprotein intakes, we used the following convention. When defining biomarker-based reference measures for nonprotein as well as nutrient density and nutrient residual, we used the first DLW observation with both the first and second repeat urinary nitrogen observations. In theory, doing so induced some correlation between repeat biomarker-based reference observations, but the DLW measurement error was so small that this correlation could be ignored in practice.

Bias in the Attenuation Factor Based on the Dietary Report Reference Instrument

For a valid reference biomarker M, the attenuation factor is expressed as $\lambda_M = \text{cov}(M, Q)/\text{var}(Q) = \text{cov}(T,Q)/\text{var}(Q)$ (18).

Thus, the biomarker-based attenuation factor λ_M is equal to the true attenuation factor λ_1. However, the attenuation factor λ_F based on the common approach with a dietary report reference instrument is given by $\lambda_F = \text{cov}(F,Q)/\text{var}(Q) = (\beta_{F1}\beta_{Q1}\sigma_T^2 + \text{cov}(r,s))/\text{var}(Q)$.

Taking into account expression 4 for the true attenuation factor λ_1, we can rewrite this expression as

$$\lambda_F = \lambda_1 \left(\beta_{F1} + \frac{1}{\beta_{Q1}} \rho_{r,s} \sqrt{\frac{\sigma_r^2 \sigma_s^2}{\sigma_T^2 \sigma_T^2}} \right). \qquad (A1)$$

Thus, the attenuation factor λ_F is generally biased. The relative bias, defined by the expression in parentheses, depends on intake-related biases in the FFQ and dietary report instrument F, reflected by slopes β_{Q1} and β_{F1}, respectively; the variances of their person-specific biases relative to variation in true intake, σ_r^2/σ_T^2 and σ_s^2/σ_T^2, respectively; and the correlation $\rho_{r,s}$ between person-specific biases. Values of slope β_{F1} less than one decrease λ_F relative to true attenuation factor λ_1, whereas positive values of

$$\rho_{r,s} \sqrt{(\sigma_r^2/\sigma_T^2)(\sigma_s^2/\sigma_T^2)},$$

as well as values of slope β_{Q1} less than one, increase λ_F.